Statistical Analysis Plan for Final Analysis

Trial title: EVA (Evaluation of Votrient in Angiosarcoma)

This Statistical Analysis Plan (SAP) describes the final analysis. This document is additional to the SAP describing the interim analysis.

Trial code: Novartis Common Code (NCC): CPZP034BDE01T
EudraCT Number: 2012-005846-39

SAP version: 1.0

SAP Date: 03.04.2020

Name and address of the responsible Biometrician:

Dr. phil. Marietta Kirchner
Institut für Medizinische Biometrie und Informatik
Im Neuenheimer Feld 130.3
D-69120 Heidelberg
Email: kirchner@imbi.uni-heidelberg.de
Phone: +49 6221 56 7784

Name and address of the representative of the Biometrician:

Dr. Anja Sander
Institut für Medizinische Biometrie und Informatik
Im Neuenheimer Feld 130.3
D-69120 Heidelberg
Email: sander@imbi.uni-heidelberg.de
Phone: +49 6221 56 5859

Name and address of the Coordinating Investigator:

Prof. Dr. med. Peter Hohenberger
Surgical Clinic,
University Hospital Mannheim, Medical Faculty Mannheim
Theodor-Kutzer-Ufer 1-3
68167 Mannheim
Email: Peter.Hohenberger@umm.de
Phone: +49 (0) 621 3832609
Fax: +49 (0) 621 3831479

Name and address of the Sponsor:

University of Heidelberg, Medical Faculty Mannheim,
represented by the chancellor of the University of Heidelberg
Central Administration of the University
Seminarstr. 2
69117 Heidelberg
Email: kanzlerin@zuv.uni-heidelberg.de
Phone: +49 (0) 6221 54-2100 & -2001
Fax: +49 (0) 6221 54-2618

Signature

Date and signature of the responsible Biometrician

Date and signature of the representative of the Biometrician

Date and signature of the Coordinating Investigator

1 Purpose of the Statistical Analysis Plan of the Final Analysis

The present Statistical Analysis Plan (SAP) describes the remaining analyses which has not already been provided at interim as it was defined that they will only be done at the final analysis. This document complements the SAP of the interim analysis. However, some sections (Section 2, 3, 5) were copied from the already existing SAP in order to have a better overview. This will be indicated in the respective sections.

The study was stopped for futility at interim. According to the study protocol, the interim analysis should have been done after having observed the primary endpoint for the first 14 patients. Actually, there was an overrunning of 12 patients so that $n = 26$ patients were included in the full analysis set (FAS) of the interim analysis. In order to guarantee that the pre-specified significance level α=0.05 is kept, the method to handle overrunning by Englert and Kieser [1] was applied. This results in the decision rule "If $s_1 \leq 17$ successes are seen out of $n_1 = 26$ patients in the first stage the trial will be stopped for futility". With 12 successes observed at interim the study was stopped. However, all analyses at interim were already provided for the $n = 26$ patients which is the total sample size of the present study. Thus, the results of these analyses are already part of the statistical report of the interim analysis and are the final results so that there is no need to repeat them here. Instead, analysis of secondary endpoints and of subgroup analyses were not part of the interim analysis and will be described here and summarized in an additional statistical report for publication purposes.

2 Objective of the Trial

Copied from the SAP of the interim analysis.

The objective of this trial is to evaluate efficacy and safety of the experimental treatment given by a combination of Pazopanib with Paclitaxel for patients with advanced or metastatic angiosarkoma (AS). Primary endpoint is progression-free survival after 6 months (observation).

3 Study design

Copied from the SAP of the interim analysis.

This is a multicentre, open, prospective, single-armed phase II trial to evaluate efficacy and safety of the experimental treatment given by a combination of Pazopanib with Paclitaxel for patients with advanced or metastatic angiosarkoma.

Additional information not written in the SAP of the interim analysis:

Tumor assessment was scheduled at baseline, every eight weeks (± 7 days) during the treatment period (consists of six cycles à four weeks (28 days), combination of Pazopanib

with Paclitaxel), 30 days after last study medication (“safety follow-up”), every twelve weeks (± 14 days) during “follow-up” until progression. According to the study protocol the follow-up period ends when the last patient has reached a follow-up period of 24 weeks.

4 Analysis sets

Full Analysis Set (FAS): Consists of all patients who were included in the trial and received the study medication at least once or in parts. The term "ITT analysis" is used for an analysis applying ITT principles to all patients of the FAS.

Note, this is the same definition as in the interim analysis.

5 Definitions of endpoints to be analysed

Copied from the SAP of the interim analysis.

Primary endpoint: The primary endpoint of the study is 6 months progression-free survival (PFS) (assessed at 182 days ± 32 days tolerance region after start of treatment, reference date is the date of imaging). Start of therapy will be counted when the patient first gets the study treatment. The assessment date which best matches the time point of 182 days will be used as the reference. If no follow-up visit matches this time interval, conclusions on the disease status at 6 months are made as follows:

- If the patient has stable disease (SD) at the next documented assessment after the time interval specified above, the status at 6 months is SD as well.
- If the patient has complete remission (CR) or partial remission (PR) at the next documented follow-up visit after the time interval specified above, the status at 6 months is counted as SD.
- If the patient has progressive disease (PD) at any follow-up visit before the time interval specified above, the status at 6 months is counted as PD.
- If neither of the above cases holds true (that is no information is available that directly allows to specify the disease status at 6 months), the status at 6 months is counted as PD.

The diagnosis of progression is based on tumor measurements according to the RECIST criteria (version 1.1, [2]) based on a predefined set of target lesions and on non-target lesions.

Secondary endpoints:

The secondary endpoints according to the study protocol are

- Overall survival (OS) defined as start of therapy (date of first study medication) until death (event).

- Response (R) according to the RECIST criteria (version 1.1, [2]). Here, best overall response (BOR) within the study period is of interest. It is defined as the best response (categories CR, PR, SD, PD) since the patient is on-treament according to the following rules: 1. If CR was documented at least once then BOR=CR, 2. If CR was never documented but PR at least once then BOR=PR, 3. If CR and PR were never documented but SD at least once then BOR=SD, 4. If CR, PR, and SD were never documented but an assessment has taken place then BOR=PD, 5. If a patient dies and reason of death is progression/ progressive disease then BOR=PD, 6. If a patient dies without progression (e.g. as reason of death) before the first assessment, “Response” is not evaluable (NE) and thus BOR=NE.
- Grading of (Serious) Adverse Events ((S)AE) according to CTCAE version 4.0 (v4.0) (these are the safety relevant endpoints)

Further secondary endpoints not specified in the study protocol but of interest for publication purposes:

- 3 months PFS (assessed at 90 days ± 30 days tolerance region after start of treatment, reference date is the date of imaging). Start of therapy will be counted when the patient first gets the study treatment. The assessment date which best matches the time point of 90 days will be used as the reference. If no follow-up visit matches this time interval, conclusions on the disease status at 3 months are made according to the rules defined for the primary endpoint.
- PFS as time-to-event endpoint (TTE) defined as start of therapy (patient first gets the study treatment) until first progression (PD) or death whatever comes first.

Subgroup analyses:

The analysis of the primary endpoint (6 months PFS) and of the secondary endpoints (3 months PFS, PFS TTE, OS and BOR) are applied to the following subgroups:

1. Cutaneous vs. visceral AS (compare Tables 8.1.1.4 and 10.2.4 of the statistical report of the interim analysis: $n = 9$ visceral, $n = 14$ cutaneous, $n = 3$ visceral and cutaneous; according to the clinical trial team, the three patients with “visceral and cutaneous” can be clearly assigned to the subgroup “cutaneous” after checking the data entries again so that two distinct subgroups will be given).
2. Primary vs. secondary AS (compare Table 8.1.1.4 of the statistical report of the interim analysis: $n = 13$ primary and $n = 13$ secondary AS)

6 Data handling

Data were checked for consistency and plausibility throughout study conduct and at the end of the study (see also data validation plan).

AE term: free text fields in the AE database are categorized into clearly defined AE terms so that identical AE have the same term assigned. MedDRA lowest level terms (LLT) are used.

In comparison to the dataset the interim analysis was based on, the following changes were made:

1. As indicated before (Section 5 “Subgroup analysis”), the grouping of patients into visceral vs. cutaneous AS was checked again by the clinical trial team due to the occurrence of unclear assignments. Four corrections were made for the present analysis: three patients assigned to “visceral and cutaneous” will be corrected to “cutaneous” (ID 07, 13, 23) and one patient assigned to “visceral” will be corrected to “cutaneous” (ID 09).
2. In addition, some open queries with respect to AE documentation were resolved. Note that these variables were not part of the interim analysis as AEs are only analysed in the final analysis.
3. For one patient (ID 25), three further follow-up visits were documented. Again, these additional visits do not influence the interim analysis (6 months PFS was already observed before) so that no changes to the interim results of the primary endpoint have to be made.

7 Statistical methods

All analyses will be applied to the FAS. All analyses have descriptive character and p-values can only be interpreted in descriptive manner.

7.1 Primary analysis

The primary analysis was part of the interim analysis. As the study was stopped at interim, these analyses are completed and will not be repeated here (compare Section 1).

7.2 Analysis of the secondary endpoints

The analysis of secondary endpoints is of descriptive nature and generally consists of summary statistics and interval estimation.

OS was defined as the time from start of therapy to death (event status = 1) or last contact (censoring, event status = 0), respectively. OS will be analysed using Kaplan-Meier curves. The median survival time will be provided alongside two-sided 95% confidence intervals (CI, if possible). Censoring events (e.g. early drop-out reasons like withdrawal, administratively censoring at the end of the study period) will be summarized using absolute and relative frequencies. Additionally, an individual listing will be provided showing information on “event” (yes/ no), “overall survival time” (time to death if event=yes or time to drop-out if event=no [days]), “reasons for drop-out” if event=no (e.g. withdrawal, lost to follow-up, other reason, end of study period).

Similarly, PFS TTE defined as the time form start of therapy to first progression (PD) or death (whatever comes first) or last contact (censoring), respectively, will be analysed by Kaplan-Meier curves with median survival time (95% CI). Again, an individual listing will be provided showing information on “event” (yes(death)/ yes(PD)/ no), “progression-free survival time” (time to death if event=yes(death) or time to PD if event=yes(PD) or time to drop-out if event=no), “reasons for drop-out” if event=no (e.g. withdrawal, lost to follow-up, other reason, end of study period).

BOR (categories PD, SD, PR, CR, NE) will be summarized by absolute and relative frequencies.

PFS rate at month 3 will be calculated as number of successes (progression-free = response status SD, PR, or CR) divided by number of patients together with exact Clopper-Pearson 95% CI.

The secondary endpoints “grading of (S)AEs according to CTCAE v4.0” are safety relevant endpoints and will be analysed by descriptive statistical methods:

1. Adverse Events (AEs – yes/no):
   Frequencies (absolute and relative) will be shown for every AE term (LLT) which occurred at least once. Here, a patient is coded “1” (yes) if the respective LLT occurred at least once and “0” (no) if it does not occur.
   If yes, grading (grade I-II (category 1), III-IV (category 2), to V (category 3)) according to the CTCAEv4.0 will be tabulated for each AE term (absolute and relative frequencies for each category). Therefore, the maximal grading for each LLT per patient is taken (this is relevant if the respective LLT is observed more than once in a patient).
2. Serious Adverse Events (SAEs - yes/no): due to the low number of SAEs (AE classified as SAE), all SAEs will be listed with description of the event, CTCAEv4.0 grading, and relation to study drug. Additionally, CTCAEv4.0 grades (grade I-II (category 1), III-IV (category 2), to V (category 3)) will be summarized by absolute and relative frequencies (over all SAES irrespective of the respective LLT).

7.3 Subgroup analyses

The analyses of PFS, OS, and BOR will be applied to the subgroups as defined in Section 5.

First, summary statistics (absolute and relative frequencies) will be given for each subgroup as corrections were applied and thus, deviations to the frequencies provided at the interim analysis are given.

Analysis of OS and PFS TTE as defined above is done by Kaplan-Meier curves stratified by subgroup (1. cutaneous vs. visceral AS, 2. primary vs. secondary AS) with median survival

time and two-sided 95% CI (if possible) provided for each subgroup. The p-value of the log-rank test for comparison of two survival curves will be reported.

Frequencies of BOR will be provided stratified by subgroup (1. cutaneous vs. visceral AS, 2. primary vs. secondary AS). Results will be displayed in contingency tables and differences in proportions is assessed by exact Pearson chi-square test (exact test preferred due to small sample size, compare Section 9). The respective test statistic with the descriptive p-value will be reported. Again, NE will be considered as a separate category and will be tabulated.

PFS rate at month 3 and at month 6 will be given for each subgroup. The point estimate for the PFS rate is calculated as number of successes (progression-free = response status SD, PR, or CR) divided by number of patients in the respective subgroup. Additionally, 95% exact Clopper-Pearson confidence intervals are provided. Difference between two proportions (1. cutaneous vs. visceral AS, 2. primary vs. secondary AS) is assessed by Barnard's exact test and the respective test statistic with the descriptive p-value will be reported. An exact test is used here due to small sample size (compare Section 9). Barnard's exact test is preferred over the more common Fisher's exact test as it is more powerful (less conservative).

8 Software

All evaluations will be carried out using the software package SAS® System 9.4 (SAS Inc., Cary/NC, USA).

9 Deviations to the study protocol

In the study protocol, the chi-square test was proposed for the analysis of differences in proportions. However, due to the small sample size (final analysis is based on $\boldsymbol{n = 26}$ patients due to futility stop at interim and not $\boldsymbol{n = 44}$ as planned in case of proceeding with the trial after interim analysis), exact tests retain the nominal alpha level and are therefore preferred even in this exploratory analysis. In the *PROC FREQ* statement in SAS, the exact statement requests the exact test as *exact barnard;* for Barnard's exact test and *exact pchi;* for exact Pearson chi-square test.

Response was defined as secondary endpoint in the study protocol without further definitions. This was specified as "best overall response" for the present analysis.

10 References

[1] S. Englert and M. Kieser (2015). Methods for proper handling of overrunning and underrunning in phase ii designs for oncology trials. Statistics in Medicine, 34:2128-2137.

[2] E.A. Eisenhauer et al. (2009). New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *European Journal of Cancer*, 45:228-247.